IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

BREAST CANCER DETECTION USING MACHINE LEARNING ALGORITHMS

Name of Author : Divyesh Shrivastava

Student of M. Tech in Computer Science & Engineering Department,

Computer Network discipline, Session 2020-2022, Oriental University, Indore, India

Abstract :

Breast cancer is a widespread issue for women globally, ranking as the second most prevalent cancer. Detecting it early is crucial for reducing mortality rates. This paper introduces a computer-aided method for breast cancer diagnosis, using a ReNet18 and Support Vector Machine (SVM). The pre-trained ReNet18 model extracts features from X-ray images, and SVM is employed for diagnosis. To enhance performance, the images undergo haze reduction for improved quality, followed by tumor segmentation using a histogram-based K-means technique to isolate the tumor region. The experiments use the Break His dataset, categorizing benign and malignant cases. The proposed model is assessed at four magnification factors (40x, 100x, 200x, 400x). Results show the model achieves 92.6% accuracy at 200x magnification. The highest specificity and precision are 93.1% and 86.5%, respectively, at 100x magnification. These findings validate the efficiency of the proposed architecture in classifying histo-pathological breast cancer cell images.

Introduction :

Cancer is characterized by the rapid and aggressive division of body cells, leading to the formation of lumps or tumors within tissues. These tumors can be categorized as either benign or malignant. Benign tumors don't spread to other parts of the body, being non-cancerous, but they can still form lumps due to increased cell growth. On the other hand, malignant tumors are cancerous, invading surrounding tissues and other parts of the body.

Among various types of cancer, female breast cancer is the most prevalent, with one in eight adult females diagnosed with it during their lifetime worldwide. Breast cancer can develop in different regions of the breast, such as ducts, lobules, or between neighboring tissues.

Early detection and diagnosis of breast cancer are crucial for increasing survival rates and minimizing false positives (misdiagnosing a cancer patient as a non-cancer patient). Several methods are employed for breast cancer diagnosis, including mammograms, computerized tomography (CT) using X-rays, MRI, and ultrasound utilizing magnetic energy and sound waves. Early-stage detection significantly improves the chances of survival and reduces the risk of false positives.

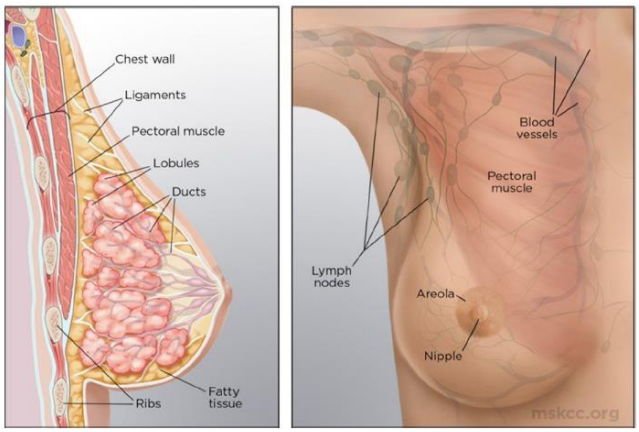

Fig 1- Breast Anatomy

The initial step in breast cancer diagnosis involves mammography, but its efficacy is hindered by the challenge of detecting cancer in women with dense breast tissue. Furthermore, the radiation emitted during mammography poses risks to both radiologists and patients. Similarly, in CT scans, radiation exposure carries the potential for genetic mutations in organs. Ultrasound, although preferred for women with dense breasts, is not deemed the most reliable diagnostic method due to its relatively high false-positive rate.

To address these challenges, a biopsy is conducted. This involves a physical examination where a sample tissue is extracted for microscopic analysis. Pathologists then scrutinize the sample under a microscope in a process known as histopathology. However, this manual examination is time-consuming, requires expertise, and is subject to the examiner's psychological fatigue, potentially leading to errors.

In recent decades, Computer-Aided Diagnosis (CAD) has emerged as the leading method for cancer detection. CAD utilizes X-ray images to segment and classify tissue samples as either benign or malignant, offering a more efficient and objective approach compared to manual histopathological examination.

Problem Statement :

In recent years, deep learning has shown remarkable success in the medical field, garnering significant attention from researchers. In this context, numerous studies have been conducted to diagnose breast cancer from histopathological images. The majority of existing work employs deep convolutional neural networks (CNN) and pre-trained models for extracting various features from histopathological images.

This study proposes an ensemble method that combines a pre-trained model with Support Vector Machines (SVM) for the detection and classification of breast tumors in histopathological images. Due to the limited availability of the dataset, a pre-trained model is utilized for feature extraction. These extracted features are then fed into an SVM for classification. Additionally, segmentation is employed to enhance accuracy and reduce the false-positive rate. This ensures that a smaller percentage of cancer images are misclassified as non-cancerous, contributing to improved diagnostic precision.

Objectives :

- To reduce ambiguity faced by K-means during clustering, spatial location is pinpointed for segmenting Regions of Interest (ROIs) from images.
- Instead of relying on a random or fixed value for K in K-means, we determine it by identifying histogram peaks. This approach enhances the precision of clustering.
- The model undergoes training and testing at various magnification factors (40x, 100x, 200x, and 400x). This not only aids in identifying the optimal magnification factor but also allows for the evaluation of model performance against existing approaches.

These strategies contribute to refining the segmentation process, improving clustering accuracy, and determining the most effective magnification factor for the model.